MAS223 Statistical Modelling and Inference Chapter 3: Likelihood

Dr Nic Freeman

School of Mathematics and Statistics, University of Sheffield

Autumn Semester, 2015

Inference

We will start to look at the idea of **statistical inference**, meaning methods of analysing data to obtain information about the processes which produced the data.

In particular, we will be looking at methods of inference based on **likelihood**, which many important aspects of statistical inference are based on.

Data

Typically we will have a set of n data values which we can think of as a vector, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

We will think of the data as being realisations of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

The random vector $\mathbf{X}$ (if continuous) will have a joint p.d.f.

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \ldots, x_n).$$

This p.d.f. will be unknown, the aim of the inference being to obtain information about it.

Independent samples

We will assume our data $x_1, x_2, \dots, x_n$ are **i.i.d. samples**.

So, it makes sense to also assume that the random variables $X_1, X_2, \dots, X_n$ are **independent and identically distributed**)

(Note here the use of capital letters for the random variables and lower case letters for the data i.e. for values they take.)

Joint p.d.f. for random sample

In this case, the joint p.d.f. $f_{\mathbf{X}}(\mathbf{x})$ will be a product of terms for each observation:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f(x_i),$$

where $f(x)$ is the common p.d.f. of the random variables $X_1, X_2, \ldots, X_n$.

Discrete case

If we have discrete random variables, then we would have a probability function instead of a p.d.f.

The theory in this case is very similar, so it is common to use the same notation in both cases:

$$f_{\mathbf{X}}(\mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

(only in this Chapter!)

Model

We assume that we already know a particular form for the joint p.d.f. of $\mathbf{X}$, usually involving some standard distribution.

However, the *parameters* of this standard distribution are unknown.

Our aim in analysing the data will be to obtain information about the (true) values of these unknown parameters.

Example

Suppose we have a biased coin, which shows heads with some (unknown) probability θ, and tails otherwise.

In this case we expect the underlying common distribution to be Bernoulli(θ).

If we toss the coin many times, we can collect data and try to guess the value of θ.

Parameters

General **parameters** are denoted by $\boldsymbol{\theta}$; we represent $\boldsymbol{\theta}$ as a vector.

Although, in any particular case the set of parameters could be a scalar (a single number θ), a matrix or some other structure.

Sometimes some of the unknown parameters will be **nuisance parameters**: their values are unknown, and we have to take account of this, but they are not what we are really interested in.

Notation

Given a model, and a particular set of parameter values $\boldsymbol{\theta}$, we write the p.d.f. (in the continuous case) for $\mathbf{X}$ as $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$.

(Again, we use the same notation in the discrete case, but it then means a probability function.)

Example

Example 26: Chemical reaction

The likelihood function

If we have a family of p.d.f.s (of p.f.s), parametrized by $\boldsymbol{\theta}$, then $f(\mathbf{x};\boldsymbol{\theta})$ defines the distribution of $\mathbf{X}$, given the parameters $\boldsymbol{\theta}$.

Once we have observed data values $\mathbf{x}$ (i.e. a realisation of $\mathbf{X}$) we can consider $f(\mathbf{x};\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

The likelihood function

When we regard $f(\mathbf{x};\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, with fixed (observed) data $\mathbf{x}$, it is called the **likelihood function**, or just the likelihood.

We write

$$L(\boldsymbol{\theta};\mathbf{x}) = f(\mathbf{x};\boldsymbol{\theta}).$$

Note that $\boldsymbol{\theta} \mapsto L(\boldsymbol{\theta};\mathbf{x})$ is not a p.d.f.

Parameter values

The likelihood $L(\boldsymbol{\theta}; \mathbf{x})$ is a function of $\boldsymbol{\theta}$.

Therefore, it is important when we calculate $L(\boldsymbol{\theta}; \mathbf{x})$ to identify the set of the possible parameter values $\boldsymbol{\theta}$, that is the domain of L.

We will denote the set of possible parameter values by Θ.

The likelihood function

Example 27: Binomial likelihood

The likelihood function, independent samples

For our model $\mathbf{X} = (X_1, \ldots, X_n)$, with parameters $\boldsymbol{\theta}$ and independent samples, we have $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$, so

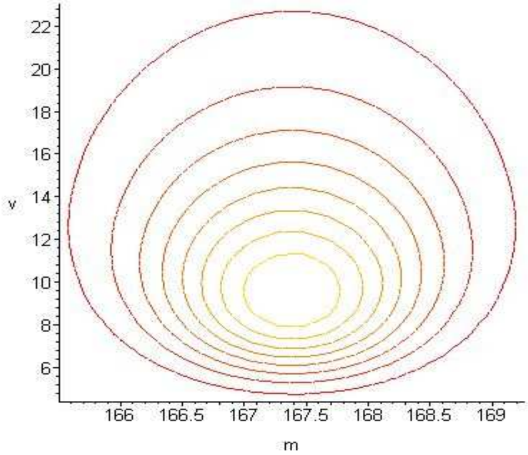
$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}).$$

Examples

Example 28: Discrete likelihood

Example 29: Chemical reaction revisited

Example 29: Contour plot



(from data 175.06, 169.13, 168.89, 165.39, 167.12, 170.83, 165.86, 164.46, 167.41, 163.79, 170.15, 167.39, 167.49, 162.24, 165.57)